(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 May 2001 (10.05.2001)

PCT

(10) International Publication Number
WO 01/33380 A1

(51) International Patent Classification[7]: G06F 15/00, 17/00, 17/21, 17/24, 15/16, 15/173

(21) International Application Number: PCT/US00/29900

(22) International Filing Date: 30 October 2000 (30.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/430,883 1 November 1999 (01.11.1999) US

(71) Applicants: GENUITY INC. [US/US]; 3 Van de Graaff Drive, Burlington, MA 01803 (US). GTE SERVICE CORPORATION [US/US]; 1255 Corporate Drive, MC SVC04A40, Irving, TX 75038 (US).

(72) Inventor: DIWAN, Arif; 1295 Phoenix Avenue, Cranston, RI 02921-1403 (US).

(74) Agents: SUCHYTA, Leonard, Charles et al.; Verizon Services Group, 600 Hidden Ridge, HQE03G13, Irving, TX 75038 (US).

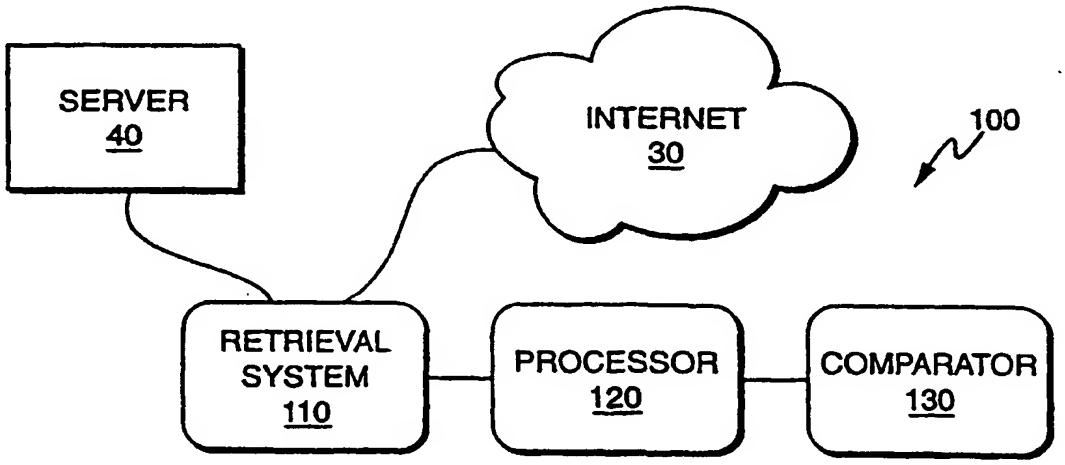
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SELF-VERIFYING REFERENCES FOR INTERNET DOCUMENTS



(57) Abstract: A system and method described herein for maintaining the integrity of electronic documents, such as web pages, which contain hyperlinks to other electronic documents. A retrieval system ((110) connected to server (40) and the internet (30) retrieves a document and calculates a value representative of the content or a portion thereof of the document referenced by the hyperlink. Comparator (130), in conjunction with processor (120), compares the changes in the calculated value so that subsequent retrievals of the referenced document may then be analyzed to verify that the contents of the documents have not been altered since the hyperlink was created.

WO 01/33380 A1

SELF-VERIFYING REFERENCES FOR INTERNET DOCUMENTS

Technical Field

This application relates to the field of document storage and retrieval, more particularly to the field of hyperlink authoring.

Background Art

The recent proliferation of Internet web sites has put a tremendous amount of information at the fingertips of anyone with a web browser. Many web sites are made up of a number of pages, each of which includes several links to other web pages, both within the web site and in other web sites, where more information can be found, or another topic can be investigated. These links, or hyperlinks, simplify navigation through the Internet and allow information to be managed in discrete chunks.

However, the fluidity with which the Internet adapts by adding, removing, or modifying web pages makes the maintenance of hyperlinks difficult. For example, a web page may include a hyperlink to a page that is subsequently deleted, making the hyperlink defective. Alternatively, the content of the referenced page may be altered, possibly in a way that affects the interpretation of the referencing page. Such an alteration may be confusing for the reader or embarrassing for the creator of the web site. Although the creator may regularly check the hyperlinks to verify that the referenced sites are still suitable, such a task can become quite time-consuming for large web sites, and is furthermore prone to overlook small changes, which may in fact have large consequences. Currently, it is difficult, if not impossible, to adequately assure the integrity of hyperlinks in a document. For example, as stated in *Proposed Technical Standards and Guidelines for Electronic Filing in the United States Courts* at http:www.cohasset.com/elec_filing/printable.html, current protocol specifically prohibits hyperlinks in electronic filings because of these problems.

Disclosure of Invention

The systems and methods described herein are useful for creating hyperlinks capable of verifying that the content of the referenced document has not been altered, e.g., is the same as the content of the document at the time the hyperlink was

created. Thus, in one aspect, disclosed herein is a hyperlink including an address of an electronic document, and a value representative of the contents of said electronic document at a predetermined time. The electronic document may be a web-based document or any other document containing a hyperlink. In certain embodiments, the value is a digitally signed value.

In another embodiment, the systems and methods described herein provide a hyperlink including means for retrieving an electronic document, and means for comparing the contents of the retrieved document to the contents of the document at a predetermined time.

In another aspect, disclosed herein is a method for creating a self-verifying hyperlink by providing an electronic document accessible at an address, determining a value representative of the contents of the electronic document, and creating a hyperlink which includes the address and the value. In certain embodiment, the method also includes digitally signing the value. In certain embodiments, creating a hyperlink includes coupling a URL address with the value.

In yet another aspect, disclosed herein is a system for monitoring the contents of electronic documents, including an address for retrieving an electronic document coupled to a value representative of the contents of a predetermined version of the electronic document, a retrieval system for obtaining a current version of the electronic document at the address, a processor for calculating a value representative of the current version of the electronic document, and a comparator for comparing the value representative of the predetermined version with the value representative of the current version to determine if the electronic document has been modified. In certain embodiments, the value representative of the predetermined version is a digitally signed value. In certain embodiments, the address is a URL address.

In another embodiment, disclosed herein is a system for verifying the contents of an electronic document, including means for locating an electronic document coupled to a value representative of the contents of the document at a predetermined time, means for retrieving the electronic document, means for generating a value representative of the contents of the retrieved document, and means for comparing the value representative of contents of the retrieved document

with the value representative of the contents of the document at a predetermined time to determine if the document has been altered since the predetermined time.

In still another aspect, disclosed herein is a method for verifying the contents of an electronic document by providing an address for retrieving an electronic document coupled to a value representative of the contents of the electronic document at a predetermined time, retrieving the electronic document from the address, determining a value for the retrieved document, and comparing the determined value with the value representative of the contents of the electronic document at the predetermined time to determine if the document has been modified since the predetermined time. In certain embodiments, providing an address includes providing a URL address, or providing an address for retrieving an electronic document coupled to a digitally signed value representative of the contents of the electronic document at a predetermined time.

In yet another aspect, disclosed herein is a web page including a hyperlink as described herein.

In another aspect, disclosed herein is system for verifying the contents of an electronic document having a retrieval system for obtaining an electronic document stored at an address, a processor for calculating a value representative of a retrieved document using a predetermined formula, and a comparator for comparing the value representative of the retrieved document with a value representative of a document previously retrieved from the address to verify that the values are identical.

In still another aspect, disclosed herein is a self-verifying hyperlink, comprising an address of an electronic document, a value representative of the contents of said electronic document at a predetermined time, and instructions for determining a value representative of the contents of the electronic document. In certain embodiments, the instructions are capable of being executed by a processor.

Brief Description of Drawings

Figure 1 illustrates a document containing hyperlinks which reference other electronic documents.

Figure 2 presents one possible structure of a self-verifying hyperlink according to the present invention.

Figure 3 depicts a computer network for verification of retrieved documents according to the present invention.

Figure 4 shows a system useful for verifying the content of retrieved documents according to the present invention.

Figure 5 illustrates a method for verifying the contents of a document retrieved using a self-verifying hyperlink according to the present invention.

Best Mode for Carrying Out the Inveniton

The description below pertains to several possible embodiments of the invention. It is understood that many variations of the systems and methods described herein may be envisioned by one skilled in the art, and such variations and improvements are intended to fall within the scope of the invention. Accordingly, the invention is not to be limited in any way by the following disclosure of certain illustrative embodiments.

Described herein are self-verifying hyperlink references and methods of using such references for ensuring that the content of a referenced document is identical to the content of the referenced document when the reference was made in the originating document containing the hyperlink reference. As illustrated in Figure 1, an electronic document 1, such as a web page, may include reference 2 to other electronic documents 3, 4, and 5, which may contain information such as text, images, charts, etc., e.g., which may supplement the content of the originating document 1. The activation of these references may retrieve the referenced documents and display them to the user, initiate downloading of the referenced document, etc. Referenced documents 3, 4, and 5 may be stored on the same server as the originating document 1, or on different servers, e.g., servers located across a network, such as the Internet. Described herein are hyperlinks, such as is schematically depicted in Figure 2, designed to permit the verification and/or validation of the content of the retrieved document, e.g., to protect against undesirable alterations in the content. As shown for the network 10 of Figure 3, such hyperlink references may be used to verify the contents of documents obtained by a client 20 from a local server 40, or from a foreign server 41 coupled to the local server 40 via the Internet 30.

As shown in Figure 2, the self-verifying hyperlink 2 may include an address portion 7 representative of the location of the referenced document, such as a URL address, and a verification portion 8 which may include a portion of the referenced document or a value representative of all or a portion of the content of the referenced document. For example, in HTML, a hyperlink may be: <A HREF http://www.refdoc.com/refdoc VERIFY=(verification portion)>, wherein VERIFY is indicative of the function used to determine the verification portion and is represented in a manner suitable for execution by a web browser or other suitable interface. Similar hyperlinks may be constructed using XML, ASN.1, or any other suitable language or encoding scheme. Changes in the content may prevent the content from being displayed or send a warning or error message to the viewer, to an administrator of the originating document, or to another appropriate person or system. In this way, changes in the content of referenced documents can be monitored to prevent a hyperlink reference from retrieving an inappropriate or undesirable document.

In one embodiment, the verification portion includes a predetermined portion of the referenced document, such as the first twenty words, characters #212-245, every sixteenth character, or any other portion as desired. When the document is retrieved, for example, by a user operating a web browser with a computer, the predetermined portion of the retrieved document is compared to the verification portion of the hyperlink.

If the two portions are identical, the retrieved document may be displayed to the user. If the two portions differ, a message may be sent to the user, for example, indicating that the content of the document has been altered, or that the document cannot be displayed. In certain embodiments, the retrieved document, although altered, may be presented to the user. Furthermore, a message may be sent to the administrator, author, or maintainer of the originating document indicating that the content of the referenced document should be verified to determine whether significant changes have been made in the content of the referenced document. Such a message may include the address of the referenced document and/or the address of the originating document.

In certain embodiments, a self-verifying hyperlink may include a value representative of all or a portion of the referenced document instead of, or in addition to, the predetermined portion. Such a value may be the result of applying a predetermined formula to the contents of all or a portion of the referenced document. Exemplary formulas that may be applied in this fashion include hashing functions, such as MD2, SHA, SHA1 and MD5, although other suitable formulas and functions will be known to those of skill in the art. Because the calculated value for a given document is difficult to predict, the use of such formulas confers the additional advantage that manipulating a document to have a different content yet identical value is rendered difficult. Thus, intentional falsification of referenced documents is severely hampered by the use of such formulas and values.

In certain embodiments, the formula used to calculate the value may be capable of distinguishing a content of a document from its format. For example, the formula may calculate a single value for a span of text whether it is stored as an Adobe Acrobat file, an HTML file, a text file, or in any other format. In this way, the value calculated by the formula better represents the content of the document, and will not indicate a change of content merely because the format of the document has been altered. Similarly, the formula may consider substantive changes, such as changes in the text, while ignoring formatting changes, such as punctuation, margins, fonts, italics, etc., which do not substantially alter the meaning of the text.

In embodiments wherein the verification portion or value is representative of a predetermined portion of a referenced document, the verification portion may be associated with or include terms indicative of the representative portion, so that the hyperlink may identify, review, and compare the predetermined portion of the referenced document. For example, in one embodiment, the verification portion may include information representative of the beginning of the representative portion and information representative of the length of the representative portion. In a different embodiment, the verification portion may be associated with information representative of the beginning of the representative portion and information representative of the end of the representative portion. Such information may be represented, for example, as XML, SGML, or HTML metatags.

8. A system for monitoring the contents of electronic documents, comprising

an address for retrieving an electronic document coupled to a value representative of the contents of a predetermined version of the electronic document,

a retrieval system for obtaining a current version of the electronic document at the address,

a processor for calculating a value representative of the current version of the electronic document, and

a comparator for comparing the value representative of the predetermined version with the value representative of the current version to determine if the electronic document has been modified.

9. A system as in claim 8, wherein the value representative of the predetermined version is a digitally signed value.

10. A system as in claim 8, wherein the address is a URL address.

11. A system for verifying the contents of an electronic document, comprising

means for locating an electronic document coupled to a value representative of the contents of the document at a predetermined time,

means for retrieving the electronic document,

means for generating a value representative of the contents of the retrieved document, and

means for comparing the value representative of contents of the retrieved document with the value representative of the contents of the document at a predetermined time to determine if the document has been altered since the predetermined time.

12. A method for verifying the contents of an electronic document, comprising

providing an address for retrieving an electronic document coupled to a value representative of the contents of the electronic document at a predetermined time,

retrieving the electronic document from the address,

determining a value for the retrieved document, and

comparing the determined value with the value representative of the contents of the electronic document at the predetermined time to determine if the document has been modified since the predetermined time.

13. A method as in claim 12, wherein providing an address includes providing a URL address.

14. A method as in claim 12, wherein providing an address includes providing an address for retrieving an electronic document coupled to a digitally signed value representative of the contents of the electronic document at a predetermined time.

15. A web page comprising a hyperlink of claim 1.

16. A system for verifying the contents of an electronic document, comprising

a retrieval system for obtaining an electronic document stored at an address,

a processor for calculating a value representative of a retrieved document using a predetermined formula, and

a comparator for comparing the value representative of the retrieved document with a value representative of a document previously retrieved from the address to verify that the values are identical.

17. A self-verifying hyperlink, comprising

an address of an electronic document,

a value representative of the contents of said electronic document at a predetermined time, and

instructions for determining a value representative of the contents of the electronic document.

18. The hyperlink of claim 17, wherein the instructions are capable of being executed by a processor.

19. A web page including a hyperlink of claim 17.

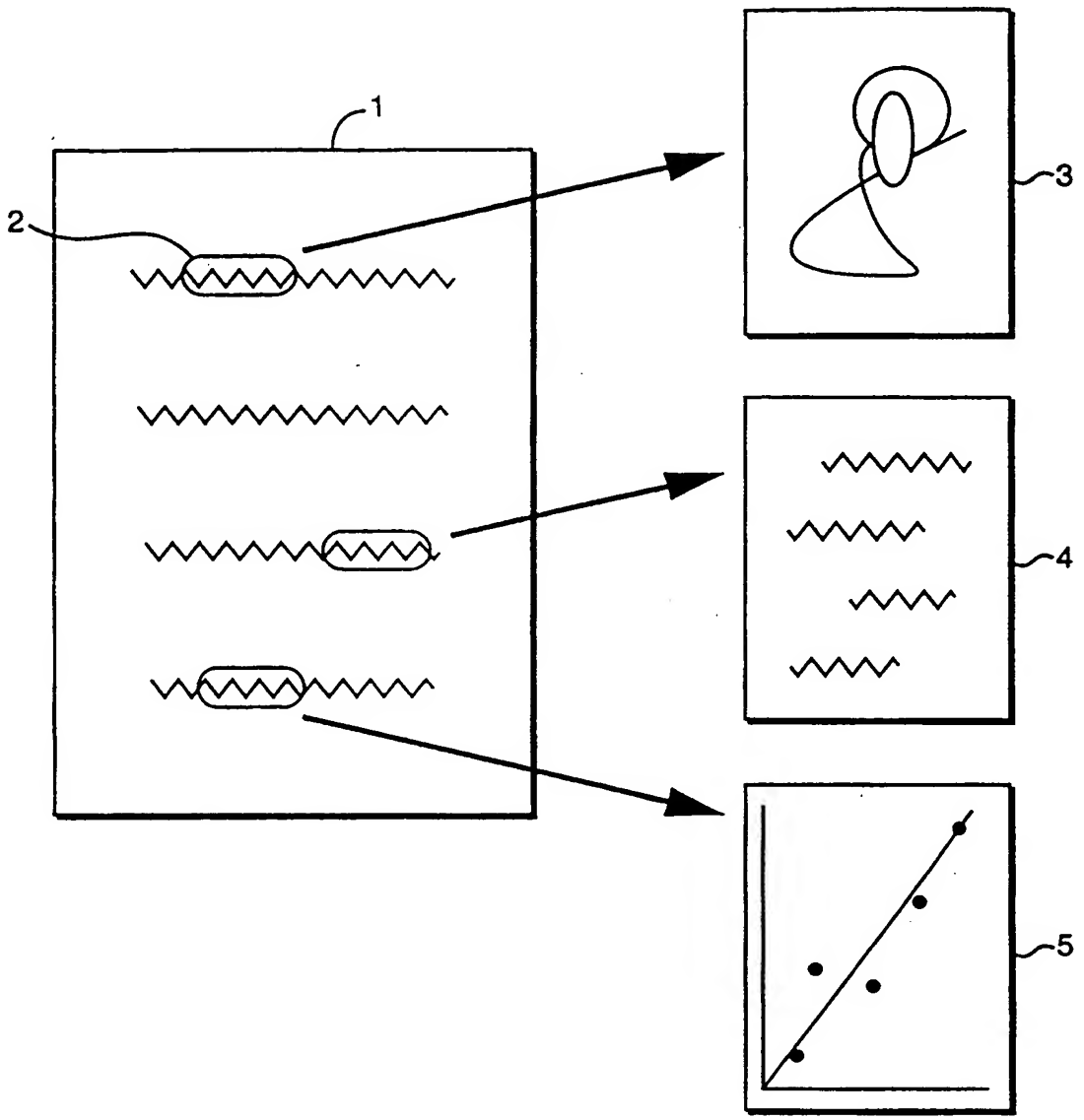


FIG. 1

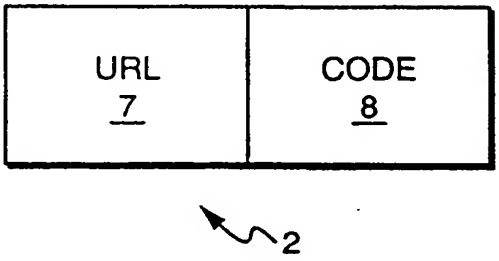


FIG. 2

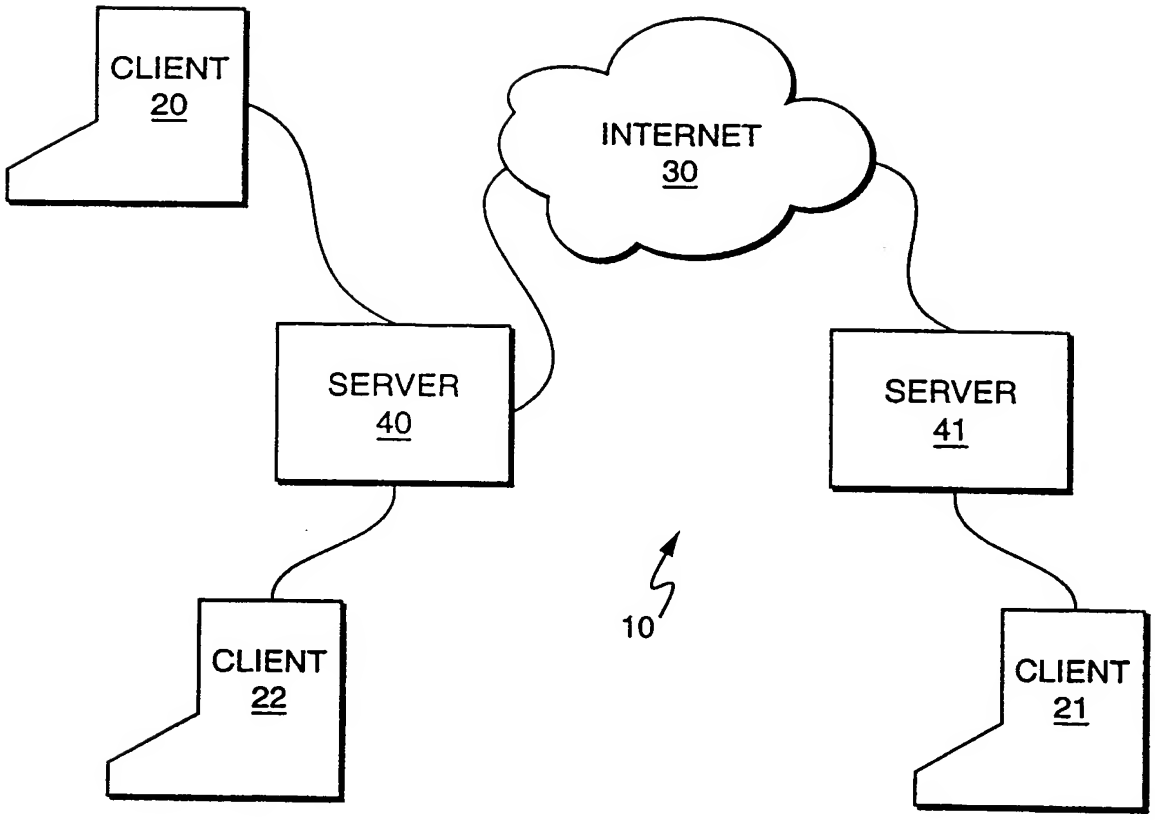

CLIENT
20
INTERNET
30
SERVER
40
SERVER
41
10
CLIENT
22
CLIENT
21

FIG. 3

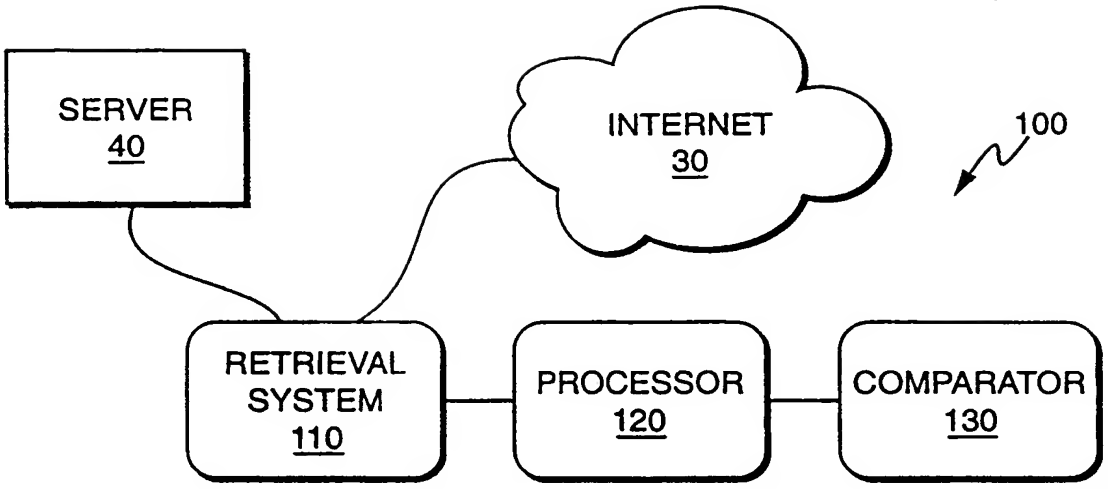

SERVER
40
INTERNET
30
100
RETRIEVAL
SYSTEM
110
PROCESSOR
120
COMPARATOR
130

FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/29900

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (7):

G06F 15/00, 17/00, 17/21, 17/24, 15/16, 15/173